Supplementary Table 1. Relevant Kids First resources, website addresses, and social accounts

Resources	Links and profiles
Cloud Credits Program	https://kidsfirstdrc.org/news/kids_first-cloud_credits
Elements of Style in Workflow Creation Course	https://nih-nichd.github.io/
GitHub Repository	https://github.com/kids-first
Listserv	https://list.nih.gov/cgi-bin/wa.exe?SUBED1=KIDSFIRST&A=1
Office Hours	https://d3b.notion.site/Kids-First-Office-Hours-a77f6aa5889c4df4ae72302f92c51aa2
Portal	https://kidsfirstdrc.org/
Social Media Profiles	Facebook - Gabriella Miller Kids First Data Resource Center Twitter - @kidsfirstDRC YouTube - @kidsfirstdataresourcecente8800
Variant Explorer	https://portal.kidsfirstdrc.org/variant
YouTube Step-by-step Guides	https://youtu.be/eVcf9Iv_pWs

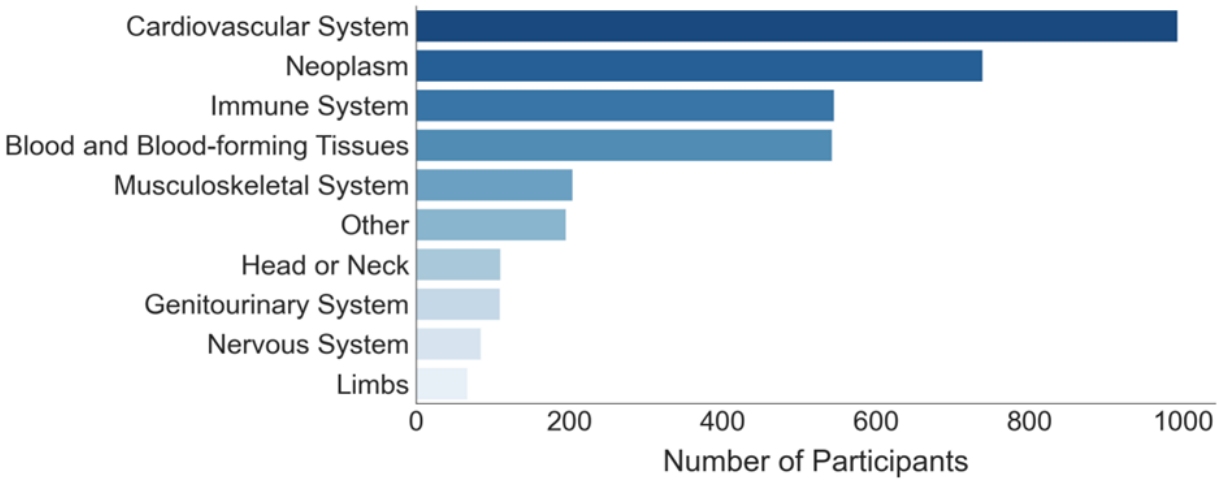
Supplementary Table 2. Funded Research Projects Using Kids First Data

Project Number*	Project Title	Contact PI	Statement of the Project's Public Health Relevance
R03CA283675	Discovering the Timing and Origins of Bone and Soft Tissue Cancers	SHLIEN, ADAM	We propose to analyze the timing and order of mutations in sarcomas, tumors that disproportionately affect children. Our premise, based on strong published and unpublished data, is that by understanding when and how a child's sarcoma develops, we will ultimately be able to predict how it will change, how it will respond to therapy and whether it will recur.
R03CA256535	Genome-wide Sequencing to Identify the Genes Responsible for Enchondromatoses and Related Malignant Tumors	SOBREIRA, NARA	The only treatment for children with Ollier disease or Maffucci syndrome is surgical, they are submitted to multiple invasive surgeries associated with long recovery periods that do not correct the deformities completely; there is no pharmacologic therapy. In addition, the risk of developing a chondrosarcoma in these individuals is ~30%, and, brain tumors, juvenile ovary and testis tumors, and vascular malignancies are also observed in these individuals with a higher frequency. We expect that our immediate results, as well as the subsequent work building on our findings, will lead to improved quality of life and medical outcomes, and reduced healthcare costs for patients with OD and MS and many others with cancers around the world.
R03CA256550	Genomic Profiling of T-cell Acute Lymphoblastic Leukemia in Children	TEACHEY, DAVID T.	Many children with T-cell acute lymphoblastic leukemia (T-ALL) relapse with current treatments and are not cured. Modern genetic tests may help us identify which patients with T-ALL are likely to relapse before they do. This project will analyze genetic testing in childhood T-ALL performed as part of a Gabrielle Miller Kids First Project and will help us predict which patients with T-ALL are at higher risk of doing poorly and allow us to give better and newer medicines to help them.
R03CA249364	Identification of Transposable Element Insertions in the Kids First Data	PARK, PETER J	Transposable elements, or "jumping genes", are genetic elements that can alter the DNA of an individual. We aim to utilize a computational method to identify such elements in the genome sequencing data generated in the Gabriella Miller Kids First Pediatric Research Program. Our analysis will identify transposable elements that may be causal for a disease phenotype.
R03CA246228	The mechanisms of somatic genome rearrangements in pediatric brain tumors	YANG, LIXING	Somatic genome rearrangements are abundant in pediatric brain tumors, but the causes are still largely unknown. We will systemically investigate the forming mechanisms of somatic rearrangements using 800 whole-genome sequenced childhood brain tumors generated by the Gabriella Miller Kids First Program. Our study will provide a better understanding of disease mechanisms, and can potentially lead to discoveries of new prognostic biomarkers and novel drug targets.

Project Number*	Project Title	Contact PI	Statement of the Project's Public Health Relevance
R03CA230366	Discovering the genetic basis of neuroblastoma initiation and progression	DISKIN, SHARON	The proposed research project is relevant to public health because we address major gaps in our understanding of the genetic basis of human neuroblastoma, an often fatal pediatric malignancy. The proposed research is highly relevant to the NIH mission of improving health outcomes as we expect that discoveries of the basic genetic mechanisms of tumor initiation and progression will lead to rational new clinical interventions.
R03CA218733	Comprehensive evaluation of new and transmitted germline variation in Ewing Sarcoma	SPECTOR, LOGAN G.	Ewing sarcoma (ES) is a rare but deadly bone cancer that occurs mainly in adolescents and young adults. Although only about 300 cases of ES occur in the United States each year, survival is low at about 60%; by identifying genes and loci associated with ES risk we may improve early detection of ES and discover possible therapeutic targets.
R03CA272955	Deep Phenotyping Children with Congenital Anomalies and Cancer Enrolled in Project:EveryChild	LUPO, PHILIP J.	One of the strongest risk factors for cancer in children and adolescents is being born with a congenital anomaly—this is true both for chromosomal abnormalities and non-chromosomal birth defects. A vital next step in this work is to assemble sufficiently large cohorts of families and children with congenital anomalies and cancer that include: 1) comprehensive phenotypic and clinical information, which can be used to ascertain syndromic characteristics, endophenotypes, and treatment outcomes; and 2) well-annotated biological samples, which can be used for wide-ranging molecular assessments. Therefore, this application will leverage the Children's Oncology Group Project:EveryChild (PEC) and the Gabriella Miller Kids First Pediatric Data Resource Center (KF-DRC) to: 1) collect extensive phenotypic and clinical data from children with congenital anomalies and cancer enrolled in PEC; and 2) integrate phenotypic and clinical data from PEC into the KF-DRC.
R03CA272952	Optimized workflows for structural variant analysis of the Kids First genomes using short and long reads	SCHATZ, MICHAEL	The Gabriella Miller Kids First Pediatric Research Long Read Pilot program aims to improve the analysis of childhood cancer and structural birth defects using new long read sequencing technologies. We will develop several workflows to improve the analysis of this data, including short-read workflows for identifying SNVs and SVs using the new Telomere-to-Telomere CHM13 reference genome, and long-read workflows for accurately identifying and comparing SVs across many samples. These workflows will then be deployed within the Kids First Data Analysis platform CAVATICA, allowing us and others to easily scale the analysis to large cohorts.

* Additional publicly available information about these projects can be found on NIH Reporter at https://reporter.nih.gov/

Supplementary Figure 1. Reported phenotypic abnormalities for a subset of participants. A subset of Kids First cancer study participants has available phenotypic data (n = 2023). This figure shows the reported categories of observed phenotypic abnormalities.



Supplementary Figure 2. Query example. A screenshot of the Kids First Portal shows an example query. Individual queries can be built and combined with Boolean operators. Dashboard visualizations update on the fly as queries are combined. The query shown selects all Kids First cancer studies (#1) and requests participants with diagnosis between age 0 and 1 (#2) with variant call files (#3).